2014 27th SIBGRAPI Conference on Graphics, Patterns and Images

Statistical Learning Approach for Robust Melanoma Screening

Michel Fornaciali, Sandra Avila, Micael Carvalho, and Eduardo Valle
University of Campinas, RECOD Lab – DCA/FEEC/UNICAMP, Campinas, SP, Brazil
{michel, sandra, micael, dovalle}@dca.fee.unicamp.br

***Abstract*—According to the American Cancer Society, one person dies of melanoma every 57 minutes, although it is the most curable type of cancer if detected early. Thus, computer-aided diagnosis for melanoma screening has been a topic of active research. Much of the existing art is based on the Bag-of-Visual-Words (BoVW) model, combined with color and texture descriptors. However, recent advances in the BoVW model, as well as the evaluation of the importance of the many different factors affecting the BoVW model were yet to be explored, thus motivating our work. We show that a new approach for melanoma screening, based upon the state-of-the-art BossaNova descriptors, shows very promising results for screening, reaching an AUC of up to 93.7%. An important contribution of this work is an evaluation of the factors that affect the performance of the two-layered BoVW model. Our results show that the low-level layer has a major impact on the accuracy of the model, but that the codebook size on the mid-level layer is also important. Those results may guide future works on melanoma screening.**

***Keywords*-melanoma skin cancer; screening; spatial pooling; BossaNova;**

I. Introduction

Melanoma is the most severe and deadly form of skin cancer [1], because of its tendency to metastasis [2]. However, it is almost always curable if detected early enough. The contrast between the prognosis of melanoma when diagnosed early and late makes early screening critical [3].

Screening, in medicine, is a strategy to detect a disease in individuals without obvious signs or symptoms, enabling earlier intervention in the hope to reduce mortality. Dermoscopy, in melanoma screening, is an in-vivo noninvasive imaging method that is useful for the early recognition of melanoma [4]. Due to the difficulty of human interpretation, automated or semi-automated analysis of dermoscopic images has become an important research area. Computer-aided diagnosis is especially valuable for isolated communities and remote regions, in which the presence of a full-time dermatologist is not feasible. Automated screening helps to select those patients who should be referred for a consultation with a specialist.

For all reasons exposed, automated or semi-automated melanoma screening has attracted the attention of both scientists and industry, and is an active research area [5]–[16]. A variety of features have been explored in literature, like color and texture descriptors combined with a binary classifier [14], or a neural network [9], [11]. Other approaches have tried to encode into the feature vectors the human-friendly 'ABCD' rule, used by dermatologists in the diagnosis of melanoma, and then to employ SVM with those features [15]. Other methods have combined color and texture features on a Bag-of-Visual-Words approach [6]–[8], one of the most successful approaches to describe the visual content of images.

Inspired by the Bag-of-Words model from textual Information Retrieval [17], where a document is represented by a set of words, the Bag-of-Visual-Words (BoVW) [18] model describes an image as a histogram of the occurrence rate of "visual words" in a "visual vocabulary" (or visual codebook) induced by quantizing the space of a low-level local descriptor (e.g., SIFT [19]). The BoVW approach has important limitations, and several alternatives to that standard scheme have been recently developed. For instance, to attenuate the effect of coding errors induced by the descriptor-space quantization, hard quantization can be replaced by a soft assignment [20] or by other coding strategies such as sparse coding [21]. To overcome the loss of spatial information, the most popular technique is the Spatial Pyramid Matching (SPM) [22]. It splits an image into hierarchical regions, which generates independent feature vectors that are concatenated into the image-level feature vector.

In this paper, we propose a statistical learning approach for robust melanoma screening. Our solution is centered on the BossaNova image representation [23], a recent extension of the BoVW approach that was shown to yield good results in many classification tasks [23], [24]. BossaNova is a statistical descriptor that replaces traditional color and texture global descriptors, leading to promising results on melanoma screening. The scheme is robust enough to forgo all ad-hoc preprocessing frequently found in the literature: there is no need to segment the image, neither to detect the lesion border, nor to detect and remove hairs, etc.

In addition to the robust screening of melanoma lesions, we study the impact of the low-level and the mid-level factors on the classification accuracy. The study is performed via a statistically sound design in order to evaluate both the significance and the relative importance of each factor in the results. No previous work in the literature perform an evaluation as broad in scope as ours, neither as rigorous in terms of statistical design.

The remainder of this paper is organized as follows. In Section II, we discuss related work on dermoscopic image analysis, highlighting the main advances on this field. In Section III, we give a detailed description of the methodology

1530-1834/14 $31.00 © 2014 IEEE
DOI 10.1109/SIBGRAPI.2014.48

used on this paper. In Section IV, we analyze our empirical results, describing the dataset and the experimental setup. Finally, Section V presents our concluding remarks and discusses future work directions.

II. Related work

Melanoma screening is an important matter on medical community, and it justifies the amount of researches on this field. This paper focus only on analysis of dermoscopic images and this section surveys the last researches on the same type of medical images.

Most of these studies tries to reproduce in computer machines the steps that dermatologists use to diagnose a melanoma. To accomplish it, some researches [9], [13], [15], [16] implement the ABCD rule [4], looking for asymmetry, border shape, color aspects and structural differences on the lesion. Others (e.g., Wadhawan et al. [5]) employ the 7-points rule [25] to classify a melanoma.

Any case have challenges that must be overcome, such as soft borders, which turn border detection into a hard problem, and the presence of veins or hair, which can impact the quality of the classification. Abbas et al. [15] deals with hair removal using derivative of Gaussian, morphological function, and fast marching techniques.

According to the literature, the process of analyzing an image of a skin lesion has three main steps: (i) identify the lesion borders (*border detection*), (ii) extract image features only inside the lesion (*feature extraction*), and (iii) compare these features with pre-calculated features of both melanoma and non-melanoma examples to decide if the skin lesion is a melanoma or not (*classification*).

Border detection can be found on [6], which implements three segmentation algorithms: ISODATA (Iterative Self-Organizing Data Analysis Technique Algorithm), fuzzy c-means and active contour without edges.

Feature extraction has been made through color and texture descriptors. The most used color descriptors are color histograms and color moments (see [5], [7], [8]). The variety for texture descriptors is vast: wavelet coefficient, Haar Wavelet, Gabor filter, Gray Level Co-Occurrence Matrix (GLCM), Active Shape Model (ASM), for instance. Those descriptors are applied in several works [5]–[8], [12]. Most studies reported that better results are reached when they combined color and texture descriptors [14].

After feature extraction, many approaches use the low-level descriptors as input for the classifier. On the other hand, some authors improve their methods by processing the low-level information before the classification step. Mid-level feature extraction aims at transforming low-level descriptors into a global and richer image representation of intermediate complexity [26]. The most popular mid-level representation is the Bag-of-Visual-Words approach (BoVW). Examples of BoVW techniques on the melanoma classification problem can be found in [6]–[8]. They work with color and texture features on the low-level and produce the mid-level by aggregating information via k-means clustering algorithm. The codebook size is small, ranging from 100 to 500 visual words [7], [8].

Traditionally, classification process have been made with Suport Vector Machines (SVM) [27], a very popular and powerful learning technique for data. Many authors applied the SVM classifier [5]–[7], [12], [13], [15]. In short, what differs one work from others is the kernel function used on the SVM and its parameters. Also, some authors employed other classification methods, such as neural networks [9], [11] or decision-trees [10], [16]. The experimental validation protocol depends on the dataset size: usually it is done with 10-fold cross-validation, but some studies like [9], [11], [14] use a leave-one-out schema.

Despite the existence of several works for melanoma classification, they are not directly compared due to the use of distinct datasets and different validation protocols among the methods. Also, there is no public melanoma dataset to facilitate different methods experimentation on same conditions.

In Section IV, Table III summarizes the main information described here and compares our results to the state-of-the-art.

III. Methodology

In this section, we describe our proposed methodology applied to the melanoma classification problem. We first provide an introduction to the BossaNova approach, the base of our framework, and we then present the Spatial Circular Pooling approach (SCP), our alternative to SPM for classifying skin lesion images. The main pipeline is shown in Figure 1.

A. BossaNova: the base of the framework

Proposed by Avila et al. [23], BossaNova is a mid-level image representation which offers a better information-preserving pooling operation based on a distance-to-codeword distribution. In this paper, we opt to use the BossaNova as the starting point of our framework due to its performance, comparing well with the state-of-the-art for several challenging datasets of image classification [23], [24].

As we have mentioned before, BossaNova is an extension of the BoVW model. It can be obtained by a succession of two steps [26]: coding and pooling. Traditionally, the coding step simply associates the image local descriptors to the closest element in the codebook, and the pooling takes the average of those codes over the entire image. Since the pooling operation compacts all the information contained in the individually encoded local descriptors into a single feature vector, that step is critical for BoW-based representations.

In general terms, the objective of pooling is to summarize the information contained in the individually encoded descriptors into a single feature vector, preserving important information while discarding irrelevant detail.

From this perspective, instead of using the classical sum- or max-pooling strategies, BossaNova introduces a density function-based pooling strategy, aggregating local spatial information about the descriptors around each codeword, preserving thus statistical information about the distribution of the features. It is done by computing a local histogram z of

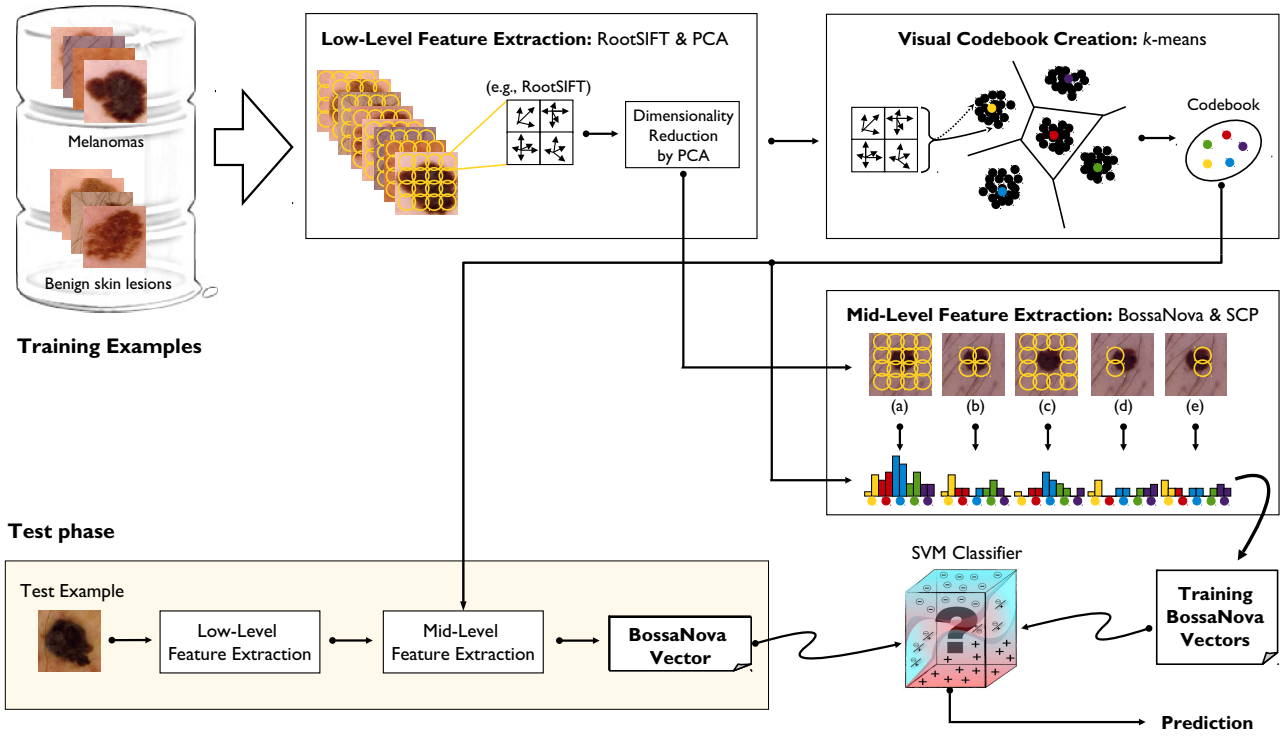

Fig. 1. The main pipeline of BossaNova. **Low-Level Feature Extraction:** RootSIFT descriptors [28], which yields superior performance than SIFT, are employed. The dimensionality of the RootSIFT is reduced from 128 to 64 by using PCA. The PCA matrix is learned over a sample of low-level features during the training phase. **Visual Codebook Learning:** During the training phase, k-means with Euclidean distance is run over a sample of one million low-level features, the final centroids are used as codewords. **Mid-Level Feature Extraction:** BossaNova descriptors creates the feature vectors for the images. The spatial pooling takes into account tone of the pyramid schemes (SPM or SCP), creating one independent feature vector for each hierarchical region of the pyramid and then concatenating them. **Decision Model Training:** During the training-phase, the BossaNova feature vectors of annotated images are employed to train a decision model using SVM (we use LIBSVM [29]). **Decision Model Prediction:** The trained model employs the BossaNova feature vectors of an image to predict on the positive (melanoma) or negative classes.

distances between the descriptors found in the image and those in the codebook. To construct the local histograms, BossaNova uses the parameters shown in Figure 2.

BossaNova vector is defined by three parameters: the number of codewords M, the number of bins B in each histogram, and the range of distances α_{MIN} and α_{MAX}. The former, α_{MAX}, avoids considering words not close enough from the center and α_{MIN} avoids the empty regions that appear around each codeword, saving space in the final descriptor. The BossaNova Z is a vector of size $M \times (B+1)$. For more details, see [23], [30].

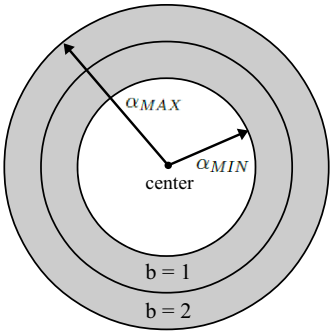

Fig. 2. Illustration of the range of distances α_{MIN} and α_{MAX} parameters of the BossaNova model. α_{MAX} avoids considering words not close enough from the center and α_{MIN} avoids the empty regions that appear around each codeword, saving space in the final descriptor. The gray area corresponds to the bounds of the histogram, local descriptors outside those bounds are ignored. Figure adapted from [23].

B. The Spatial Circular Pooling Approach

The new Spatial Circular Pooling (SCP) introduced in this paper is a type of spatial pooling addressed specially for the skin lesion classification problem. It enriches the BossaNova representation by adding spatial information about the image descriptors distribution around the skin lesion. Also, the SCP can be extended to other BoVW based techniques in a straightforward manner.

Typically, dermoscopic images are concentric, that is, the lesion is centered on the image and it occupies about 50% of the image area. Furthermore, many researches segment the

lesion before extracting image features. Here, we propose a new, fast and easy way to extract the lesion without need to segment the image. The method is explained in Figure 3 (top row): we draw a circular region with radius R to capture 50% of the image area (see Equation 1), we consider five sampling vectors composed by (a) the whole image, (b) the outer and (c) the inner regions and (d) the left and (e) the right sides of the lesion. The schemas (a)-(c) try to evaluate the impact of lesion segmentation over the classification, and the schemas (d)-(e) try to identify asymmetrical borders, which is a relevant criteria according to the ABCD rule of dermoscopy [4].

$$R = L/\sqrt{2\pi}, \tag{1}$$

where R is the radius of the circle used on the SCP approach and L is the size of the square skin lesion image. In our experiments, the images are resized to 316 $\times$ 316 pixels, so the radius R is 126.1 pixels, starting at the center of the image.

In order to evaluate our new spatial pooling strategy, we compare it with the most popular spatial pooling approach: the Spatial Pyramid Matching (SPM) proposed by Lazebnik et al. [22]. As shown in Figure 3 (bottom row), in this technique the image is divided on regular grids creating a pyramid of pooled features. The descriptors are organized into vectors, one for each grid cell. In the illustration, the image is divided on a 1×1+2×2 schema, that is, the image is divided into a 1×1 grid (the whole image – (a)), and into a 2×2 grid (four spatial cells – (b) to (e)).

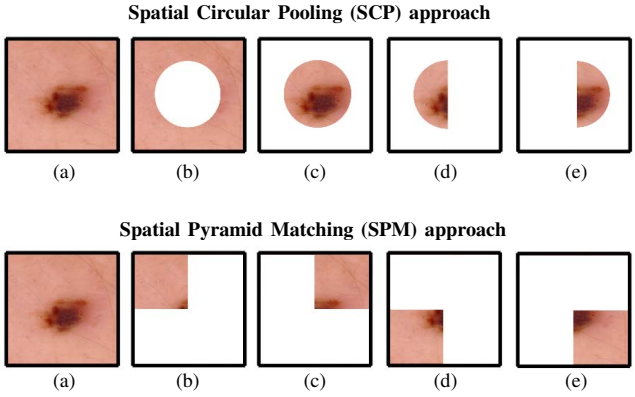

Fig. 3. Comparison between the *Spatial Circular Pooling* (SCP) (top row) and the Spatial Pyramid Matching (SPM) approach [22] (bottom row), contrasted in our evaluation of the factors affecting the model accuracy. While SCP tends to emphasize the contrast between the center and the border of the image, SPM tends to emphasize the contrast between the quadrants of the image. While we expected the center–border contrast to be very important (because it corresponds to the rules of dermoscopy image analysis), our results show that SPM and SCP perform equally.

For a fair comparison, we have chosen the 1×1+2×2 schema for the SPM pooling because it has the same dimensionality of our approach, that is, the feature vectors will both have the same size (five regions). This comparison aims to identify the informative power of each spatial pooling strategy, that is, given a feature vector of the same size, we will detect which approach preserves more information about the image. It is straightforward to note that the more information the feature vector has, the better is the classification.

IV. Experiments

In this section, we describe the dataset, the experimental setup, as well as the experimental results.

A. Dataset

The dataset is a third-party development and was created by the Department of Medical Informatics, RWTH Aachen University. It is composed of 747 dermoscopic images with resolution of 512×512 pixels, of which 187 images are melanomas and 560 images are benign skin lesions. The dataset is not publicly available, but can be download after a license agreement is signed[1]. A few example images are shown in Figure 4.

Researchers tend to use dermoscopic instead of clinical images (captured by a common camera under non-controlled conditions) to detect melanomas automatically due to the better quality, generally highlighting the lesion and its color and texture structures.

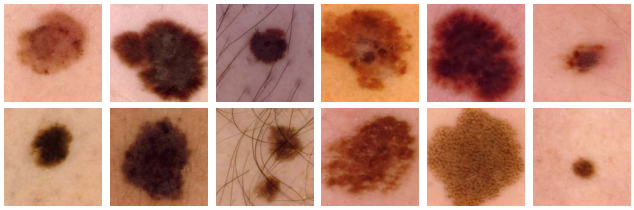

Fig. 4. Extracts of skin lesions. Melanomas (top row) and benign skin lesions (bottom row) appear very similar, making the task very challenging.

B. Experimental setup

Our experiments aim to identify the power of statistical descriptors on the melanoma classification problem, as well as to clarify the impact of low-level and mid-level over the classification.

Images are resized to an area of 100K pixels, if larger. In the low-level feature extraction, we extract RootSIFT descriptors [28], [31]. The dimensionality of the RootSIFT is reduced from 128 to 64 by using principal component analysis (PCA). To learn the codebook, we apply the k-means clustering algorithm with Euclidean distance over one million randomly sampled descriptors. In the mid-level feature extraction, we use the BossaNova representation which has shown good results in recognition tasks [23]. We incorporate spatial information using our proposed Spatial Circular Pooling (SCP) scheme or the Spatial Pyramid Matching (SPM) approach. Classification is performed by support vector machine (SVM) classifiers using LIBSVM library [29]. We employ the grid search to find the best parameters for SVM. The area under the receiver

[1]IRMA datasets – http://ganymed.imib.rwth-aachen.de/irma/datasets

operating characteristic (ROC) curve, the AUC, is used to measure the classification performance.

We apply a 5-fold cross-validation, generating nearly 600 images for training and 150 images for testing on each fold.

In order to provide common ground for future comparisons, the code used in this paper is publicly available[2].

Evaluated parameters: We have chosen six attributes of the framework, analyzing their contribution for the final result. The low-level attributes are step and size values of the Root-SIFT extraction. The mid-level attributes are the codebook size, the number of bins, the maximum and minimum values for α parameter and the pooling schema. The setup values are detailed on Table I. We will discuss these impacts on the next section. Detailed results are provided on our website.

TABLE I
PARAMETERS OF THE FRACTIONAL FACTORIAL EXPERIMENT. EACH LINE SHOWS A PARAMETER OF THE EXPERIMENTAL SETUP, TESTED WITH HIGH AND LOW VALUES, IN ORDER TO IDENTIFY ITS IMPACTS ON THE CLASSIFICATION RESULT.

	Parameter		Value
Low-level	**step**	*small*	8
		big	24
	scale min	*min*	12
		max	24
	scale max	*min*	64
		max	128
	#scale	*few*	2
		many	4
Mid-level	**#bins**	*few*	2
		many	4
	α	*tight*	[0.6, 1.6]
		loose	[0.2, 2.0]
	codebook	*small*	1024
		big	2048
	pooling	SPM [22]	1×1+2×2
		SCP (ours)	(see Fig.3)

C. Results and discussion

In order to test the significance of the parameters and their interactions, a fractional factorial ANOVA, with a block design using the folds as blocks, was employed. To create an evaluation measure for the ANOVA analysis, we have used the logit(AUC) since the AUC is a highly non-linear measure, and the ANOVA method assumes linearity in the data. After that, we removed the average of each fold, in order to make comparable the 5 replications between them. The study of significance is done on this final variable. Due to the aliasing of the fractional design, we have considered only second-order interactions, yet none of them came out significative.

The statistical results are shown on Table II, which presents that the residuals contain most of the information about variability in the classification. This indicates that the choice of parameters is not obvious and no parameters combination improves systematically the results of the folds. We conclude, therefore, that every researcher should pay particular attention to the random choice of the images to compose the folds for the experiments. Each fold must be balanced between melanoma and non-melanoma images and other factors that can be bypassed without contaminating the results, like hair presence or not, good and bad illumination and lesion size.

Despite of the residuals, the main effects that came significative were for step and min scale (p-value < 0.001), as well as the number of bins and the codebook size. These parameters have strong influence on the classification, because their p-values indicate that the occurrence of the same results due to randomness is practically impossible. We conclude, then, that these parameters are good clues to be explored when constructing a BoVW-based method for melanoma classification.

The statistical analysis still shows that the low-level has bigger impact over the classification than the mid-level. It can be proved by the column 'Sum of squares': note that the low-level concentrates higher values for this parameter. This shows that in order to construct a good melanoma classifier, we should pay attention on the feature extraction step. Although less informative than the low-level, the mid-level also plays an important role on the classification. Remarking that the effect of the codebook size was very significative (p-value = 6.6×10^{-4}), this shows that the choice of the codebook size significantly improves the predictive power of the model. In addition, an analysis of the ANOVA table shows that the step choice was, arguably, the most influential factor (largest partition of the mean square variation), reinforcing the relevance of that factor on improving the classification model.

To conclude this section, we will provide a deeper analysis of the spatial pooling strategies. The average AUC for both SPM and SCP approaches is 93.7% and the ANOVA analysis did not show statistical differences between each approach. This shows that both approaches offer the same results and the proposed pooling schema (SCP) is so good as that one introduced by Lazebnik et al. [22] (SPM) for melanoma classification problem. It is important to note, and it is possible to see in Figure 3, bottom row, that the SPM pooling schema captures information about the asymmetry of the lesion by grouping the descriptors in 4 regions. The SCP schema, on the other hand, just captures information about the borders and the center of the lesion, clustering descriptors in two groups: one with descriptor belonging to the lesion and other with descriptors that do not belong to the lesion. We can also conclude that the segmentation proposed in Figure 3, top row, frames (d) and (e) are not sufficient to capture the whole asymmetry of the lesion on the SCP approach. These evidences suggest that in the melanoma classification problem the investigation of the lesion asymmetry is so important as its segmentation.

Comparison with the state-of-the-art: This section aims to compare our results with that ones presented on the state-of-the-art. It is extremely important to note that the comparison is done just in an illustrative way since the results are not directly comparable due to the use of distinct datasets for validation. Unfortunately, we cannot implement other researchers methods due to non-disclosure and copyright agreements.

[2]https://sites.google.com/site/robustmelanomascreening/

TABLE II

PARTIAL VIEW OF THE ANOVA TABLE. WE OMIT THE SECOND-ORDER INTERACTIONS SINCE NONE OF THEM WERE SIGNIFICANT. ON THE OTHER HAND, ALL MAIN EFFECTS WERE SIGNIFICANT. THE CHOICES OF STEP AND SCALE FOR THE LOW-LEVEL AND THE CHOICE OF NUMBER OF BINS AND CODEBOOK SIZE FOR THE MID-LEVEL EXPLAIN MOST OF THE NON-RANDOM VARIATION, AS SEEN IN THE SUM OF SQUARES COLUMN.

Level	Parameter	Degrees of freedom	Sum of squares	Mean square	F value	p-value	
Low-level	**step**	1	40.66	40.66	110.76	$< 2.00 \times 10^{-16}$	***
	scale_min	1	9.64	9.64	26.26	5.45×10^{-7}	***
	scale_max	1	1.60	1.60	4.35	3.79×10^{-2}	*
Mid-level	**#bins**	1	4.85	4.85	13.21	3.29×10^{-4}	***
	codebook	1	4.35	4.35	11.85	6.60×10^{-4}	***

Significance codes: *** p-value < 0.001; ** p-value < 0.01; * p-value < 0.05

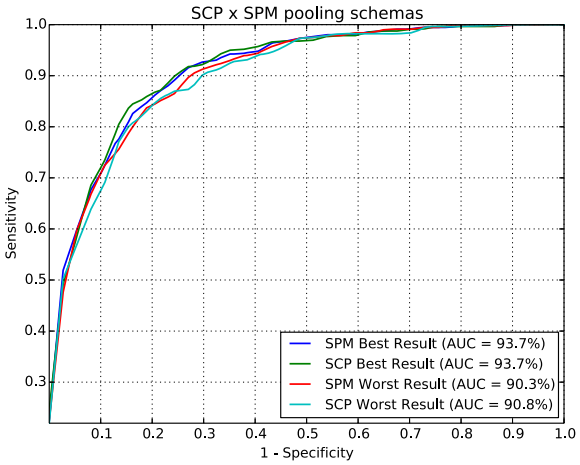

Fig. 5. ROC curves of the best and worst average AUCs for SCP and SPM pooling approaches. Note that despite our best average AUCs are 93.7% for both SCP and SPM, SCP has a worst case better than SPM.

Nevertheless, the comparison is important to prove that our results are promising and aligned to the state-of-the-art.

Figure 5 shows the ROC curves of the best and worst results for SCP and SPM pooling approaches. Our best average AUCs are 93.7% for both SCP and SPM, proving that our pooling schema is suitable for the melanoma classification problem.

Table III resumes the information present on the state-of-the-art section. It also shows important aspects about each study, like the dataset size, the proportion between positive (melanoma) and negative (non-melanoma) images and the evaluation criteria: the AUC or the accuracy. We will compare our results only with studies that use AUC as evaluation measure, since we consider it more informative than the accuracy, due the aggregation of information about the sensitivity and the specificity of the methods, while the accuracy favors the true positive and true negative results, and masks the false positive and false negative ones.

Our method is directly better than [6], [7], [9], and [15]. Also, it should be mentioned that Iyatomi et al. [9] has border detection and feature selection, forcing a non-natural improvement of the method.

In our experiments, we used the whole dataset, without removing difficult cases, like images with poor quality, excessive presence of hair or if the lesion is not whole fitted on the image. We also do not detect lesion borders, remove hair, improve the contrast between melanoma and non-melanoma skin nor do any other ad-hoc preprocessing to benefit the classifier, proving the robustness of our method.

It is also important to note that none of the state-of-the-art works reported standard deviation of their results, preventing a deeper analysis of the actual behavior of the proposed methods. We, however, presented our deviations by showing the residuals of the ANOVA analysis.

V. CONCLUSION

This paper presents a novel approach for the melanoma classification problem based on the state-of-the-art BossaNova descriptor, which considerably extends the BoVW model and has shown good results on several tasks. We were able to create a robust melanoma screening technique, which forgoes all ad-hoc preprocessing frequently found in literature like image segmentation, lesion border detection, or hair and vein detection and removal. The comparison with the state-of-the-art is very difficult due to the fact there are no standard datasets, and the contacted authors refused to share either the data or code leading to their publications. Nevertheless, our results are very promising, with an AUC of up to 93.7%.

In addition to proposing a new technique for the robust screening of melanoma lesions, we have studied the impact of the low-level and the mid-level factors on the classification accuracy. The study was performed with statistically sound design in order to evaluate both the significance and the relative importance of each factor in the results. No previous work in the literature perform an evaluation as broad in scope as ours, neither as rigorous in terms of statistical design. We were able to show that the scale and the step of the low-level features are among the most-influential parameters.

Future works include the continued optimization of both low-level and mid-level layers of the representation, with special attention to the factors pointed as most influential in our study. In addition, we want to implement the proposed method into mobile devices, enabling screening programs on remote communities.

TABLE III

RESULTS REPORTED IN THE LITERATURE. BECAUSE THERE IS NO STANDARD DATASET NEITHER PROTOCOL TO ALLOW DIRECT COMPARISON, THE RESULTS REPORTED BY LITERATURE ARE NOT DIRECTLY COMPARABLE, AND ARE SHOWN HERE FOR ILLUSTRATION AND FOR COMPLETENESS SAKE. NOTE AS WELL, THAT SOME AUTHORS EMPLOY AUC WHILE OTHERS EMPLOY THE ACCURACY (ACC) AS METRIC.

Ref.	Authors	Method	Dataset #pos/#neg	AUC (%)	ACC (%)
[5]	Wadhawan et al.	Color histogram; Haar wavelet; SVM	110/237	*	76.4
[6]	Wadhawan et al.	Haar wavelet; SVM	388/912	91.4	*
[7]	Situ et al.	Color histogram; Gabor filter; BoVW; SVM	30/70	82.2	*
[8]	Barata et al.	Color histogram; Gabor filter; BoVW; k-NN	25/151	**	**
[9]	Iyatomi et al.	Color and texture descriptors; Neural network	198/1060	92.8	*
[11]	Mikos et al.	GLCM; Neural network	42/88	*	69.5
[12]	Doukas et al.	ASM; SVM	800/2200	*	85-90
[14]	Marques et al.	Color and texture descriptors; *	17/146	*	79.1
[15]	Abbas et al.	ABCD rule-based features; SVM	60/60	88.0	*
–	Our proposal	RootSIFT; BossaNova & SCP; SVM	187/560	93.7	–

AUC: area under the ROC curve — ACC: accuracy — *It was not reported by the authors in the original paper
**Uses Sensitivity (93%) and Specificity (85%), as evaluation measure

ACKNOWLEDGMENTS

This research was partially supported by FAPESP, CAPES, CNPq and Project "Capacitação em Tecnologia de Informação" financed by Samsung Eletrônica da Amazônia Ltda., using resources provided by the Informatics Law no. 8.248/91. We would like to thank CENAPAD (Project 533) and Microsoft Azure for the infrastructure used on the experiments. Also, we are grateful to Prof. Dr. Thomas M. Deserno, from the Department of Medical Informatics, RWTH Aachen University, for kindly sending the skin cancer images to our research. Many thanks to Ramon Pires for helpful discussions.

REFERENCES

[1] "Skin Cancer," http://www.skincancer.org/. Last acessed: July 04, 2013.

[2] "American Cancer Society. Cancer Facts & Figures 2013," last acessed: July 04, 2013.

[3] A. Jerant, J. Johnson, C. Sheridan, and T. Caffrey, "Early detection and treatment of skin cancer," *American Family Physician*, vol. 62, no. 2, pp. 357–368, 2000.

[4] F. Nachbar, W. Stolz, T. Merkle, A. Cognetta, T. Vogt, M. Landthaler, P. Bilek, O. Braun-Falco, and G. Plewig, "The ABCD rule of dermatoscopy: High prospective value in the diagnosis of doubtful melanocytic skin lesions," *Journal of the American Academy of Dermatology*, vol. 30, no. 4, pp. 551–559, 1994.

[5] T. Wadhawan, N. Situ, H. Rui, K. Lancaster, X. Yuan, and G. Zouridakis, "Implementation of the 7-point checklist for melanoma detection on smart handheld devices," in *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2011, pp. 3180–3183.

[6] T. Wadhawan, N. Situ, K. Lancaster, X. Yuan, and G. Zouridakis, "Skinscan: A portable library for melanoma detection on handheld devices," in *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2011, pp. 133–136.

[7] N. Situ, X. Yuan, J. Chen, and G. Zouridakis, "Malignant melanoma detection by bag-of-features classification," in *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* 2008, pp. 3110–3113.

[8] C. Barata, J. Marques, and T. Mendonça, "Bag-of-features classification model for the diagnose of melanoma in dermoscopy images using color and texture descriptors," in *International Conference on Image Analysis and Recognition (ICIAR)*, 2013, pp. 547–555.

[9] H. Iyatomi, H. Oka, M. Celebi, M. Hashimoto, M. Hagiwara, M. Tanaka, and K. Ogawa, "An improved internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm," *Computerized Medical Imaging and Graphics*, vol. 32, no. 7, pp. 566–579, 2008.

[10] G. Di Leo, A. Paolillo, P. Sommella, and G. Fabbrocini, "Automatic diagnosis of melanoma: A software system based on the 7-point checklist," in *43rd Hawaii International Conference on System Sciences (HICSS)*, 2010, pp. 1–10.

[11] E. Mikos, I. Sioulas, K. Sidiropoulos, and D. Cavouras, "An android-based pattern recognition application for the characterization of epidermal melanoma," in *Workshop on Biomedical Instrumentation and Related Engineering and Physical Sciences (BIOMEP)*, 2012.

[12] C. Doukas, P. Stagkopoulos, C. Kiranoudis, and I. Maglogiannis, "Automated skin lesion assessment using mobile technologies and cloud platforms," in *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2012, pp. 2444–2447.

[13] M. Mete and N. Sirakov, "Dermoscopic diagnosis of melanoma in a 4d space constructed by active contour extracted features," *Computerized Medical Imaging and Graphics*, vol. 36, no. 7, pp. 572–579, 2012.

[14] J. Marques, C. Barata, and T. Mendonca, "On the role of texture and color in the classification of dermoscopy images," in *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2012, pp. 4402–4405.

[15] Q. Abbas, M. Celebi, I. Garcia, and W. Ahmad, "Melanoma recognition framework based on expert definition of ABCD for dermoscopic images," *Skin Research and Technology*, vol. 19, no. 1, pp. 93–102, 2012.

[16] G. Capdehourat, A. Corez, A. Bazzano, R. Alonso, and P. Musé, "Toward a combined tool to assist dermatologists in melanoma detection from dermoscopic images of pigmented skin lesions," *Pattern Recognition Letters*, vol. 32, no. 16, pp. 2187–2196, 2011.

[17] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*, 1st ed. Addison Wesley, 1999.

[18] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *International Conference on Computer Vision (ICCV)*, vol. 2, 2003.

[19] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2004.

[20] J. van Gemert, C. Veenman, A. Smeulders, and J.-M. Geusebroek, "Visual word ambiguity," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 32, pp. 1271–1283, 2010.

[21] Y. Boureau, F. Bach, Y. LeCun, and J. Ponce, "Learning mid-level features for recognition," in *Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 2559–2566.

[22] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Computer Vision and Pattern Recognition*, 2006, pp. 2169–2178.

[23] S. Avila, N. Thome, M. Cord, E. Valle, and A. de A. Araújo, "Pooling in image representation: the visual codeword point of view," *Computer Vision and Image Understanding (CVIU)*, vol. 117, no. 5, pp. 453–465, 2013.

[24] ——, "BossaNova at ImageCLEF 2012 Flickr Photo Annotation Task," in *Working Notes of the Conference and Labs of the Evaluation Forum (CLEF)*, 2012.

[25] G. Argenziano, G. Fabbrocini, P. Carli, V. D. Giorgi, E. Sammarco, and M. Delfino, "Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. comparison of the abcd rule of dermatoscopy and a new 7-point checklist based on pattern analysis," *Arch Dermatology*, vol. 134, no. 12, pp. 1563–1570, 1998.

[26] Y. Boureau, F. Bach, Y. LeCun, and J. Ponce, "Learning mid-level features for recognition," in *Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 2559–2566.

[27] V. N. Vapnik, *The nature of statistical learning theory*. Springer-Verlag New York, Inc., 1995.

[28] R. Arandjelovic and A. Zisserman, "Three things everyone should know to improve object retrieval," in *Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 2911–2918.

[29] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, pp. 27:1–27:27, 2011.

[30] S. Avila, N. Thome, M. Cord, E. Valle, and A. de A. Araújo, "BOSSA: Extended bow formalism for image classification," in *International Conference on Image Processing (ICIP)*, 2011, pp. 2909–2912.

[31] A. Vedaldi and B. Fulkerson, "VLFeat - An open and portable library of computer vision algorithms," in *ACM International Conference on Multimedia*, 2010, pp. 1469–1472.